Moss Expressing Promoting Regions

The invention relates to isolated nucleic acid molecules promoting expression of polypeptides in genetically modified eukaryotic host cells.

The expression of proteinaceous substances (proteins, peptides, polypeptides, fragments thereof, as well as posttranslationally modified forms of these molecules are hereinafter referred to as "polypeptides" (synonymously used together with "protein", e.g. in the example part) in genetically modified cells is a major source for providing preparations of such often rare and valuable substances. For expressing such polypeptides in genetically modified host cells, the presence of a DNA region is necessary which positively controls ("activates", "promotes") this expression. Promoters are important examples for such regions allowing RNA polymerases to bind to the DNA for initiating transcription into mRNA (Watson et al., "Recombinant DNA" (1992), Chapter I.1 and 2).

Mosses have gained increasing attention as useful objects for research for plant physiology and development, since their simple nature (mosses are situated at the base of higher-plant-evolution) provides insights into the complex biology of higher plants. The simple morphology of mosses and the advantageous culturing possibilities has made them popular model organisms for studies of plant physiology and developmental biology: Moss species may be cultured without difficulty under controlled conditions, using in vitro techniques including axenic culture, not only in petri dishes, but also in liquid culture e.g. in bioreactors. The haploid gametophyte can be grown photoautotrophically in sterile culture and easily observed at the cellular level.

Another major advantage of mosses is their transformation capacity: Despite numerous studies, the ratio of targeted integration events in plants hardly reaches 10^{-4}, which prevents the general use of gene targeting approaches for plant functional genomics. In contrast to all other plants having been tested so far, integration of homologous DNA sequences in the genome of

mosses (especially the established moss model organisms such as Physcomitrella patens (for a review of its molecular genetics: Reski, 1999)) occurs predominantly at targeted locations by homologous recombination. Transformation of mosses is usually and easily performed via PEG-mediated uptake of plasmid DNA by protoplasts, DNA transfer by microprojectile bombardment, electroporation and microinjection (Cove et al., 1997). Depending on the design of the transforming construct predominantly random or targeted integration occurs.

Despite the use of mosses as scientific tools for plant physiology research, the use of mosses for producing recombinant heterologous polypeptides in moss cells has been rather limited so far, although efficient production methods have become available (e.g. culturing protonema moss tissue as described in EP 1 206 561 A).

A major limitation of transformation technologies in eukaryotic host cells, especially in animal cells or cells of higher plants, has always been the lack of an efficient promoter for high constitutive expression of foreign genes in such transgenic host cells. The cauliflower mosaic virus (CaMV) 35S promoter has been widely used for this purpose in a number of plant transformation systems (see e.g. WO 01/25456 A), however, the CaMV 35S promoter has shown low activity in some plant species (specially monocots, such as rice (McElroy et al., 1991,)). For monocot transformation the rice actin 1 5' region has been used for heterologous expression of proteins (McElroy et al., 1991,). Nevertheless, the continuing need to provide novel expression promoting means for the expression of recombinant (foreign) polypeptides in genetically modified eukaryotic host cells still exists.

For mosses, especially for Physcomitrella patens, up to now, no homologous (in this case homologous is defined as: moss derived) suitable nucleus derived expression promoters or other nucleus derived expression promoting sequences have been published so far (Holtdorf et al., 2002). Researchers have therefore used heterologous (in this case heterologous is defined as: not moss derived) promoters for the expression of selection marker genes

and other genes of interest. However, only a few of such promoters have been reported to function reliably in certain mosses (e.g. the CaMV 35S-promoter; summarised in Holtdorf et al., 2002; CaMV 35S-promoter does not work in certain other species (Zeidler et al., 1999); TET-promoter (reviewed in Reski (1998)). Therefore, other means for genetically manipulating mosses have been developed in the art, e.g. gene-trap and enhancer trap systems (Hiwatashi et al., 2001; however, also using (a shortened version of the) CaMV 35S promoter; the authors showed in transient expression experiments that also thist shortened version of the 35 S promoter was functioning as a weak promoter; in fact, this paper relates to the expression of a reporter gene in enhancer-trap strains but does not reveal any correlation of this expression to any regulatory element of mosses).

Whereas in the above mentioned research in mosses using homologous recombination the use of heterologous promoters is necessary (and therefore homologous promoters are not needed, moreover they are in most cases not useful), the need for a suitable moss derived expression promoting means for industrially using mosses for the production of recombinant polypeptides or for the overexpression of homologous polypeptides is present and yet unsolved. Such expression promoting means should allow a stable and constitutive expression under the applied culturing conditions and should preferably enable a comparable or even higher expression performance as the CaMV 35S promoter.

Therefore, the present invention provides an isolated nucleic acid molecule encoding a moss expression promoting region (MEPR), i.e. an expression promoting region from a wild type moss. With the present invention moss derived expression regions (i.e. nucleus derived regions originating from wild type mosses) are provided which allow a constitutive expression in genetically modified host cells, especially mosses, thereby addressing the needs for such tools raised in the prior art (Holtdorf et al., 2002; Schaefer et al., 2002).

An essential feature of the MEPRs according to the present invention is also that the expression promoting activity of the MEPRs is at least 30 %, preferably at least 50 %, of the expres-

sion promoting activity of a working heterologous promoter in the specific host cell (e.g. CaMV 35S for the expression of a recombinant polypeptide in Physcomitrella patens), because moss promoters which do not have such an expression promoting activity cannot be properly used for solving the objects of the present invention and are therefore not regarded as MERPs.

The MEPRs according to the present invention are therefore isolated from the nucleus of wild type mosses, i.e. mosses which have not been genetically modified by the introduction of promoters from non-moss species (e.g. promoters of higher plants or (plant) pathogens, such as the CaMV 35S promoter, or the TET promoter). It is also clear that MEPRs with minor sequence variation (e.g. exchange of 1, 2, 3, 4 or 5 bases in regions which do not negatively affect (abolish) the expression promoting activity), which may occur e.g. due to natural strain sequence variability or due to events during isolation of the MEPRs are also regarded as MEPRs according to the present invention. Methods for analysing the expression promoting activity or for analysing the effect of such minor sequence variation on this activity are available to the skilled man (e.g. by comparison with the known CaMV 35S constructs) and also described in the example section below.

According to the present invention MEPRs promoting expression which is not sphorophyte specific, are defined as constitutive MEPRs, preferably MEPRs promote expression in gametophyte derived cells, more preferably MEPRs promote expression in protonema cells.

According to the present invention constitutive expression is preferably defined as the expression of a protein resulting in detectable amounts of this protein under liquid culture conditions generally used for photoautotrophically grown mosses, e.g. flask cultures, bioreactor cultures (EP 1 206 561 A), conditions used for the transient expression system described beneath. Therefore, constitutive expression has to be given for the MERPs according to the present invention preferably without the need of specific culturing additives, preferably also without the need of added sugars, phytohormones or mixtures of such sub-

stances in the culture medium. The constitutive expression has to be performed in a steady mode; yet it can be transient.

The terms "moss" or "mosses" as used in the present specification encompasses all bryophytes (hepatics or liverworts, hornworts and mosses). Characteristic for mosses is their heteromorphic *Generationswechsel*, the alternation of two generations which are distinct from each other in terms of nuclear DNA amounts and morphology. The diploid sporophyte is photosynthetically active only in its youth and requires supply from the dominating, green, haploid gametophyte. The gametophyte exists in two morphologically distinct forms: the juvenile gametophyte, called protonema and the adult gametophyte, called gametophor. In contrast to the protonema, the adult gametophyte (gametophore) bears the sex organs.

In the context of the presented invention transient expression is defined as introduction of an episomal nucleic acid-based construct (e.g. MEPRs and gene of interest) as descibed below into a moss protoplast and causing or allowing transient expression from the vector that results preferably in turn to the secretion of extracellular protein into the medium. Protoplasts are derived from moss cells, preferably, from gametophytic cells, more preferably from protonema cells.

Although the MEPRs according to the present invention may be taken from any moss species, the MEPRs are preferably isolated from common model moss species. The MEPRs are therefore preferably isolated from Physcomitrella, Funaria, Sphagnum, Ceratodon, Marchantia and Sphaerocarpos, especially of Physcomitrella patens, Funaria hygrometrica and Marchantia polymorpha.

Suitable MEPRs according to the present invention are selected from the Seq. ID Nos. 1 to 27 or expression promoting fragments thereof. An "expression promoting fragment" is a fragment of an MEPR which has an expression promoting activity of the MEPRs of at least 30 %, preferably at least 50 %, of the expression promoting activity of a working heterologous promoter in the specific host cell (e.g. CaMV 35S for the expression of a recombinant polypeptide in Physcomitrella patens).

The MEPRs according to the present invention may comprise specific regions, such as a promoter region ("promoter"), 5'untranslated regions ("5'-UTRs"), 5'-introns or 3'-UTRs. For some MEPRs, expression promoting fragments exist which only contain the 5'-intron. Usually the promoter is always active alone as an expression promoting fragment. Therefore, the MEPR according to the present invention preferably comprises a moss promoter and preferably a 5'-UTR region and/or a 5'-intron and/or a 3'-UTR .

Although it is often sufficient, if a certain constitutive expression is reached, it is in many cases preferred to achieve a high expression rate, especially for industrially producing recombinant polypeptides. Most of the MEPRs according to the present invention have proven to allow significantly higher expression rates for a given recombinant polypeptide than the CaMV 35S promoter, especially in homologous systems (e.g. a Physcomitrella MEPR for expression of a polypeptide in Physcomitrella). Therefore, preferred MEPRs according to the present invention have an expression promoting activity being at least equal to the expression promoting activity of cauliflower mosaic virus (CaMV) 35S promoter, especially, but not limited, in the moss species from which the MEPR was isolated. Even more preferred MEPRs have an expression promoting activity being at least 200 %, preferably being at least 500%, especially being at least 1000 %, of the expression promoting activity of cauliflower mosaic virus (CaMV) 35S promoter, especially, but not limited, in the moss species from which the MEPR was isolated.

The isolated nucleic acid molecules according to the present invention are preferably used to transform a specific host cell for producing a recombinant transgenic polypeptide, preferably, but not limited to, in an industrial scale. Therefore the nucleic acid molecule is provided as a suitable vector allowing transformation and expression of the transgene in the host cell. Among the possibility that an MEPR according to the present invention is used for replacing a natural promoter in mosses, thereby bringing the expression of a homologous moss polypeptide under the control of a MPER being located at a position in the genome of the moss, where it is normally not present in wild

type strains, the prevalent industrial applicability of the present MEPRs is the control of expression of a heterologous ("foreign") gene in a production host cell, specifically a plant cell, especially a moss cell. Therefore, the nucleic acid molecule according to the present invention further comprises a coding region for a recombinant polypeptide product, said coding region being under the control of the MEPR.

It is also advantageous, if the isolated nucleic acid molecules according to the present invention further comprises a selection marker and/or further regions necessary for enabling the appropriate transformation method chosen (see e.g. Cove et al., 1997; Schaefer, 2002). For example, if targeted integration is preferred, the nucleic acid molecule according to the present invention should further comprise sequences which are homologous to genomic sequences of the species to be transformed. Thus, allowing targeted integration of the isolated nucleic acid molecule via homologous recombination into the genome of the species to be transformed.

Moreover, the isolated nucleic acid molecules according to the present invention can be used for screening and defining consensus sequences for expression promoting regions. Finding and screening for such consensus sequences (regions, boxes) which are important and/or essential for expression promoting activity is a valuable asset in recombinant DNA technology, especially with respect to industrial biotechnology using mosses.

According to another aspect, the present invention also relates to a process for the expression of a recombinant polypeptide product in an eukaryotic host cell comprising the following steps:

- providing a recombinant DNA cloning vehicle comprising an isolated nucleic acid molecule encoding an MEPR according to the present invention and optionally a coding region for said recombinant polypeptide product, said coding sequence being under the control of the MEPR of said nucleic acid molecule in said host,
- transforming said eukaryotic host cell which does not naturally harbour said coding sequence in a way that it is under the control of said MEPR,

- culturing the transformed eukaryotic host cell in a suitable culture medium,
- allowing expression of said recombinant polypeptide and
- isolating the expressed recombinant polypeptide.

As mentioned above, MEPRs according to the present invention in principle have the capability to achieve constitutive expression in various cell types, the eukaryotic host cell is preferably selected from plant cells, preferably moss cells, especially Physcomitrella patens cells.

A system which is specifically preferred for the present invention is the culturing in moss protonema cultures (protonema moss tissue). In doing so the method described in the EP 1 206 561 A and the preferred embodiments thereof are explicitly incorporated by reference herein and are immediately applicable to the present invention.

The constitutive expression of the polypeptide with the means according to the present invention is possible without the need for various additives in the culture medium, specifically without additives for specific differentiation or promoting different tissue growth. Therefore, besides electrolytes, selection agents and medium stabilisers, the culture medium preferably does not contain any further additives for cell supply. The culture medium for stably transformed plants is preferably free from added sugars, phytohormones or mixtures thereof. The culture medium for transiently transformed protoplasts is preferably free from added phytohormones.

Preferred moss cells are moss cells of the group Physcomitrella, Funaria, Sphagnum, Ceratodon, Marchantia and Sphaerocarpos, especially in protonema cultures.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding an MEPR for industrially producing a polypeptide, especially for providing recombinant cells producing said polypeptide. The industrial production allows a large scale preparation of a given polypeptide of interest in bioreactors, e.g. in gram amounts or even

higher (commercial yields). This in contrast to the production sufficient for research use (mg amounts)or analytical purposes (µg amounts), which may, of course also be performed by the present invention. In transient expression systems, protein amounts sufficient for such analytical purposes can easily be obtained with the present DNA molecules.

Accordingly, the present invention also encompasses the use of an isolated nucleic acid molecule encoding a MEPR for expression of a moss polypeptide, the expression of said moss polypeptide being not naturally controlled by said MEPR, especially for providing recombinant moss cells expressing said polypeptide. This use may be reduced to practice both, for research purposes and for industrial scale production of moss polypeptides.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding a MEPR for expression of proteins involved in specific posttranslational modifications (e.g. glycosyltransferases), especially for providing recombinant moss cells expressing polypeptides with posttranslational modifications normally not existing or normally existing in another ratio in untransformed moss cells.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding a MEPR for expression of proteins involved in metabolic pathways, especially for providing recombinant moss cells altered in their contents of metabolites e.g. secondary metabolites.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding a MEPR for expression of antisense molecules, siRNA molecules or ribozymes especially for providing recombinant moss cells with reduced amounts of specific proteins resulting in altered phenotypes e.g. morphologically, biochemically.

According to another preferred aspect, the present invention also relates to the use of an isolated nucleic acid molecule encoding an MEPR according to the present invention for recombinant expression of postranslationally modifying proteins,

especially for the production of posttranslationally modified proteins. With such a technology, it is possible to produce proteins which are specifically modified postranslationally (differently than in the native host cell, thereby enabling e.g. plant cells or moss cultures to allow the production of proteins with e.g. mammal or even human glycosylation patterns. Examples wherein such techniques are applied with specific glycosyltransferases are described e.g. in WO 00/49153 A and WO 01/64901 A.

Another preferred use of the isolated nucleic acid molecule encoding an MEPR according to the present invention relates to the in vitro expression of recombinant proteins. The technique of in vitro translation allows a more controlled production of the recombinant product without the need to accept the uncertainties being connected with host cells.

Another preferred use of the nucleic acid molecule according to the present invention is their use for recombinant expression of metabolism modifying proteins, e.g. proteins which modify the (posttranslational) modification of a translated amino acid chain (see e.g. Berlin et al, 1994).

The present invention is further illustrated by the following examples and the figures, yet without being restricted thereto.

Figures:

Fig.1 ß-tubulin genes in Physcomitrella patens,
Fig.2 Analysis of expression promoting regions of ß-tubulins in Physcomitrella patens,
Fig.3 Analysis of expression promoting regions of Pptub 1 by transient transformation of rhVEGF constructs,
Fig.4 Analysis of expression promoting regions of Pptub 2 by transient transformation of rhVEGF constructs,
Fig.5 Analysis of expression promoting regions of Pptub 3 by transient transformation of rhVEGF constructs,
Fig.6 Analysis of expression promoting regions of Pptub 4 by transient transformation of rhVEGF constructs,
Fig.7 Genomic structure of Physcomitrella patens actin genes,
Fig.8 Comparison of the expression activity of the different 5

'actin regions,
Fig.9 Ppact1 constructs,
Fig.10 Ppact 5 constructs,
Fig.11 Ppact 7 constructs,
Fig.12 Pp act3::vegf constructs,
Fig.13 Ppact1 promoter:5' intron substitutions,
Fig.14 Ppact1 promoter:vegf deletion constructs,
Fig.15 Ppact3 promoter:vegf deletion constructs,
Fig.16 Ppact5 promoter:vegf deletion constructs,
Fig.17 Ppact7 promoter:vegf deletion constructs,
Fig.18 Actin genes in various moss species, and
Fig.19 Comparison of promoter sequences of homologous actin genes from Physcomitrella patens and Funaria hygrometrica

Material and Methods

Plant material

Physcomitrella patens (Hedw.) B.S.G. has been characterised previously (Reski et al. 1994)). It is a subculture of strain 16/14 which was collected by H.L.K. Whitehouse in Gransden Wood, Huntingdonshire, UK and was propagated by Engel (1968; Am J Bot 55, 438-446).

Standard culture conditions

Plants were grown axenically under sterile conditions in plain inorganic liquid modified Knop medium (1000 mg/l $Ca(NO_3)_2$ x 4 H_2O 250 mg/l KCl, 250 mg/l KH_2PO_4, 250 mg/l $MgSO_4$ x 7 H_2O and 12.5 mg/l $FeSO_4$ X 7 H_2O; pH 5.8 (Reski and Abel (1985) Planta 165, 354-358). Plants were grown in 500 ml Erlenmeyer flasks containing 200 ml of culture medium or on 9 cm Petri dishes with solidified Knop medium (10g/l agar). Flasks were shaken on a Certomat R shaker (B.Braun Biotech International, Germany) set at 120 rpm. Conditions in the growth chamber were 25 +/- 3°C and a light-dark regime of 16:8 h. Cultures were illuminated from above by two fluorescent tubes (Osram L 58 W/25) providing 35 micromol/$m^{-2}s^{-1}$. Subculturing of liquid cultures was done once a week by disintegration using an Ultra-Turrax homogenizer (IKA, Staufen, Germany) and inoculation of two new 500 ml Erlenmeyer

flasks containing 100 ml fresh Knop medium. Additionally, cultures were filtered 3 or 4 days after disintegration and were transferred into fresh Knop medium.

Bioreactor cultures were grown in Knop medium or in 1/10 Knop medium, respectively, in stirred tank glass bioreactors (Aplikon, Schiedam, The Netherlands) with a working volume of 5 liters (as described in Hohe and Reski, Plant Sci. 2002, 163, 69-74). Stirring was performed with a marine impeller running with a speed of 500 rpm, the cultures were aerated with 0.3 vvm [(aeration volume)/(medium volume)/min] air. The culture temperature of 25°C in the vessel was controlled by a double jacket cooling system. Light intensity was 50 micromol/$m^{-2}s^{-1}$ provided by fluorescent tubes (Osram L 8W/25) with a light/dark rhythm of 16/8 h. The pH-value in the culutures (pH 6.5 - 7.0) was not adjusted.

Protoplast Isolation

Different protocols for the isolation of protoplasts (Grimsley et al. 1977; Schaefer et al. 1991; Rother et al. 1994; Zeidler et al. 1999; Hohe and Reski 2002; Schaefer 2001) have been described for Physcomitrella patens. For the work presented herein, a modification/combination of the previously described methods was used:

Moss tissue was cultivated for 7 days in Knop medium with reduced (10%) $Ca(NO_3)_2$ content. Cultures were filtered 3 or 4 days after disintegration and were transferred into fresh Knop medium with reduced (10%) $Ca(NO_3)_2$ content. After filtration the moss protonemata were preincubated in 0.5 M mannitol. After 30 min, 4% Driselase (Sigma, Deisenhofen, Germany) was added to the suspension. Driselase was dissolved in 0.5 M mannitol (pH 5.6-5.8), centrifuged at 3600 rpm for 10 min and sterilised by passage through a 0.22 µm filter (Millex GP, Millipore Corporation, USA). The suspension, containing 1% Driselase (final concentration), was incubated in the dark at RT and agitated gently (best yields of protoplasts were achieved after 2 hours of incubation). The suspension was passed through sieves (Wilson, CLF, Germany) with pore sizes of 100 micrometer and 50 micromet-

er. The suspension was centrifuged in sterile centrifuge tubes and protoplasts were sedimented at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro, Germany). Protoplasts were gently resuspended in W5 medium (125 mM $CaCl_2$ x 2 H_2O; 137 mM NaCl; 5.5 mM glucose; 10 mM KCl; pH 5.6; 660-680 mOsm; sterile filtered; Menczel et al. 1981). The suspension was centrifuged again at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro, Germany). Protoplasts were gently resuspended in W5 medium. For counting protoplasts a small volume of the suspension was transferred to a Fuchs-Rosenthal-chamber.

Transient Transformation

Different protocols for transformation (Schaefer et al. 1991; Reutter and Reski 1996, Schaefer 2001) have been described for Physcomitrella patens. For the work presented herein, a modification/combination of the previously described methods was used:

For transformation protoplasts were incubated on ice in the dark for 30 minutes. Subsequently, protoplasts were sedimented by centrifugation at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). Protoplasts were resuspended in 3M medium (15 mM $CaCl_2$ x 2 H_2O; 0.1% MES; 0.48 M mannitol; pH 5.6; 540 mOsm; sterile filtered, Schaefer et al. (1991) Mol Gen Genet 226, 418-424) at a concentration of 1.2 x 10^6 protoplasts/ml. 250 microliter of this protoplast suspension were dispensed into a new sterile centrifuge tube, 50 microliter DNA solution (column purified DNA in H_2O (Qiagen, Hilden, Germany, Hilden, Germany); 10-100 microliter optimal DNA amount of 60 microgram was added and finally 250 microliter PEG-solution (40% PEG 4000; 0.4 M mannitol; 0.1 M $Ca(NO_3)_2$; pH 6 after autoclaving) was added. The suspension was immediately but gently mixed and then incubated for 6 min at RT with occasional gentle mixing. The suspension was diluted progressively by adding 1, 2, 3 and 4 ml of 3M medium. The suspension was centrifuged at 20°C for 10 minutes at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). The pellet was resuspended in 400 microliters 3M medium. Cultivation of transformed protoplasts was performed in 48 well plates (Cellstar, greiner bio-one, Frickenhausen, Ger-

many).

Transient transformations were incubated in dim light (4.6 micromols-1m-2) at 25°C. Samples were taken after 24h and 48h, respectively, by carefully replacing half of the medium (200 microliters) by fresh medium. The medium was not replaced completely since the protoplasts have to be kept in liquid. The removed medium (including recombinant protein) was stored at -20° C. The 48h samples were measured in an ELISA.

Stable transformation

Different protocols for transformation (Schaefer et al. 1991; Reutter and Reski 1996, Protocol Schaefer 2001) have been described for Physcomitrella patens. For the work presented herein, a modification/combination of the previously described methods was used:

For transformation protoplasts were incubated on ice in the dark for 30 minutes. Subsequently, protoplasts were sedimented by centrifugation at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). Protoplasts were resuspended in 3M medium (15 mM $CaCl_2$ x 2 H_2O; 0.1% MES; 0.48 M mannitol; pH 5.6; 540 mOsm; sterile filtered, Schaefer et al. (1991) Mol Gen Genet 226, 418-424) at a concentration of 1.2 x 10^6 protoplasts/ml. 250 microliter of this protoplast suspension were dispensed into a new sterile centrifuge tube, 50 microliter DNA solution (column purified DNA in H_2O (Qiagen, Hilden, Germany, Hilden, Germany); 10-100 microliter optimal DNA amount of 60 microgram was added and finally 250 microliter PEG-solution (40% PEG 4000; 0.4 M mannitol; 0.1 M $Ca(NO_3)_2$; pH 6 after autoclaving) was added. The suspension was immediately but gently mixed and then incubated for 6 min at RT with occasional gentle mixing. The suspension was diluted progressively by adding 1, 2, 3 and 4 ml of 3M medium. The suspension was centrifuged at 20°C for 10 minutes at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). The pellet was re-suspended in 3 ml regeneration medium. Selection procedure was performed as described by Strepp et al. (1998).

ELISA

Recombinant VEGF121 expressed by transient transformed moss protoplasts was quantified by ELISA (R&D Systems, Wiesbaden, Germany). The ELISA was performed according to the instructions of the manufacturer. The samples were diluted for quantification.

Bacterial strains and cloning vectors

For all cloning and propagation experiments Escherichia coli strain Top10 (Invitrogen, Karlsruhe, Germany) was used. For cloning of DNA-fragments pCR2.1-TOPO (Invitrogen, Karlsruhe, Germany), pCR4-TOPO (Invitrogen, Karlsruhe, Germany), pZErO-2 (Invitrogen, Karlsruhe, Germany) or pRT101 (Töpfet et al. (1987), NAR, 15, p5890) were used as vectors.

Genomic DNA: preparation, digestion, ligation

Physcomitrella patens genomic DNA was isolated from 13 days old protonemata following the CTAB protocol (Schlink and Reski, 2002).

Genomic DNA (3-5 micrograms) was digested with 30 units of various restriction endonucleases (e.g. BamHI, EcoRI, HindIII, KpnI, NcoI, NdeI, PaeI, PagI, XbaI; all MBI Fermentas, St. Leon-Rot, Germany) in a total volume of 30 microliters for two hours at 37°C, using one endonuclease per digest. Digested DNA was purified using PCR Purification Columns (Qiagen, Hilden, Germany), following the suppliers manual (30 microliters digest + 200 microliters buffer PB). Elution was done in 50 microliters Elution Buffer (EB; Qiagen, Hilden, Germany). Prior further treatment, 10 microliters of the eluate were analysed on an agarose gel (0,5%).

The remaining DNA was religated with 5 units T4 Ligase (MBI Fermentas, St. Leon-Rot, Germany) in a total volume of 300 microliters for two hours at RT and additional two days at 4°C. Prior addition of the enzyme ligation mixtures were put for five minutes at 50°C and then on ice, in order to melt sticky end basepairing. After ethanol precipitation with 0,3 M Na-acetat (pH 4.8) and two washes with 70% ethanol the religated DNA was

resuspended in 200 microliters EB . One to three microliters of this religated genomic DNA were used for I-PCR.

RNA Preparation

Physcomitrella patens total RNA was prepared by grinding tissue under liquid nitrogen and by the usage of E.Z.N.A. Plant RNA Kit (PeqLab) or RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the suppliers manuals. Total RNAs were gel analysed, quantified (OD260), and stored at -20°C or -80°C, respectively.

DNase treatment and First Strand cDNA Synthesis

1 microgram of total RNAs was DNase (GIBCO BRL) digested in a total volume of 11 microliters, following the suppliers manual. 4,5 microliters of this DNase treated total RNA (~400ng) was used with Oligo dT(12-18) primers and SUPERSCRIPT II RNase H Reverse Transcriptase (GIBCO BRL) to prepare first strand cDNA, following the suppliers manual. The resulting cDNA was 10 times diluted with sterile ddH_2O and stored at -20°C.

PCR in general

If not indicated in particular PCRs were done with Advantage cDNA Polymerase Mix (BD Biosciences Clontech, Heidelberg, Germany). For all other PCR-approaches the following DNA polymerases were used: Taq recombinant polymerase (MBI Fermentas, St. Leon-Rot, Germany), Pfu native polymerase (MBI Fermentas, St. Leon-Rot, Germany), Platinum Pfx DNA polymerase (Invitrogen, Karlsruhe, Germany) or TripleMaster PCR System (Eppendorf, Hamburg, Germany). Licenced Thermo-cyclers were Mastercycler gradient (Eppendorf, Hamburg, Germany). All primers were synthesised by MWG Biotech AG, Ebersberg, Germany. For PCR product purification or gel elution GFX PCR DNA and Gel Band Purification Kit (Amersham Bioscience, Freiburg, Germany) was used, following the suppliers manual.

Construction and Cloning of Recombinant Plasmids

Conventional molecular biology protocols were essentially as de-

scribed by Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Inverse PCR (I-PCR) & nested PCR

I-PCR was done with 0.25 microliters Advantage cDNA Polymerase Mix and buffer (including 3,5 mM $Mg(OAc)_2$, both BD Biosciences Clontech, Heidelberg, Germany), 0.2 mM each primer, 0.2 mM dNTPs and one to three microliters of genomic religations (see above) in a toatl volume of 25 microliters. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96°C, 10 seconds initially 67°C (touchdown: -0.15°C/cycle) and 10 minutes 68°C as a second step, with 35 to 40 repetitions, followed by a terminal step of 20 minutes at 68°C and cooling to 4°C at the end of the program. PCR products were eluted from agarose gels. Elution was done in 30 microliters. Eluted PCR products were either cloned directly in TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany) or used as template for reconfirmation in nested PCRs. In the latter case gel eluted, nested PCR products were cloned in TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany). Cycling conditions for nested PCRs were: an initial step of 1 minutes at 96°C, then 20 seconds 94°C, 10 seconds 56°C and 4 minutes 68°C as a second step, with 25 repetitions, followed by a terminal step of 10 minutes at 68°C.

Generation of pRT101new for cloning of amplified promoter fragments

pRT101p21 (Gorr 1999) was reamplified with Pfu native polymerase (MBI Fermentas, St. Leon-Rot, Germany) using primer 320 and 321 (for this and all subsequent primers see Table 1). Primer 320 (forward) starts at the 2nd codon (5´-(atg)aac...) of the VEGF signal peptide. Primer 321 (reverse) starts in the middle of the HincII site within the multiple cloning site in front of the 35S promoter (5´-gac...). An additional XhoI site was introduced with primer 321. Religation of the PCR product resulted in loss of the 35S promoter and the reconstitution of a HincII site. The sequence of the VEGF gene was verified by sequencing. This new vector was called pRT101new and used for cloning of expres-

sion promoting regions via the XhoI or HincII site, respectively, in front of the reporter gene.

Sequencing

All sequencing reactions were performed by SEQLAB Sequence Laboratories, Göttingen, Germany

Software

Sci Ed Central, Clone Manager Suite were used for primer design, pairwaise and multiple sequence alignments. Lasergene, DNASTAR (Version 5) Megalign and SeqMan was used for analysing sequencing data. Homology searches were carried out by BLAST 2 (Altschul et al., 1997).

EXAMPLES

The present invention is illustrated by four examples for moss expression promoting regions: first, the isolation and analysis of various members of a family of tubulin expression promoting regions of Physcomitrella patens. In the second example expression promoting regions for the actin gene family from a variety of different mosses are provided. The third and fourth example deals with ubiquitin expression promoting regions and with RBCS expression promoting regions.

EXAMPLE 1: Cloning and analysis of Physcomitrella patens ß-tubulin genes and their expression promoting regions.

Overview

In order to get ß-tubulin (tub) regulatory/promoter sequences from Physcomitrella patens (Pp) in a first step coding sequences of ß-tubulin homologues were isolated by polymerase chain reaction (PCR). Therefor an alignment of all nine published ß-tubulin genomic sequences from Arabidopsis thaliana (Attub 1-9) were used to design primers within highly conserved coding regions (8F, 9F and 10R; for this and all subsequent primers see Table 1). In addition, sequence information of public EST data from Physcomitrella patens were used, but only three did show

homologies to ß-tubulins. One of which was used to design a gene-specific primer (F7) upstream of the predicted coding region. Sequence comparison of all cloned PCR products, generated with the primers mentioned and EST data lead to 3 groups of clones with identical DNA within but differences between groups, mainly, but not exclusively, due to differences within introns. This ß-tubulin orthologues were named Pptub 1, Pptub 2 and Pptub 3, respectively.

Furthermore, since during the running project, more EST data were available (more than 50000 new entries in NCBI/dbEST with beginning of 2002), a detailed analysis of all 121 Physcomitrella patens ESTs with high similarity to ß-tubulin lead to three additional new upstream and three downstream groups of ESTs, being identical within a group but neither identical to any other group nor to Pptub 1-3. PCR with primers derived from predicted noncoding upstream and downstream regions (see below) from each new group and permuting all primer combinations helped to correlate corresponding upstream and downstream groups to a particular locus, named Pptub 4, Pptub 5 and Pptub 6, respectively. Both, genomic and cDNA amplificates of all three new loci were cloned and sequenced, raising the number of ß-tubulin orthologues in Physcomitrella patens to six.

Pptub 1 to 4 (in contrast to Pptub 5 and 6) are much more frequently represented in EST databases. Corresponding cDNA libraries were produced using RNA mainly from protonema and young gametophore. So, for this four genes only, based on the gained sequence data, an inverse PCR approach (I-PCR) was performed in order to walk into flanking genomic regions.

Pptub 1

As already mentioned in a first step, Taq (MBI Fermentas, St. Leon-Rot, Germany) PCR fragments from two independent PCRs on Physcomitrella patens genomic DNA using primers 8F and 10R were cloned. One clone (2-1) and two clones (8-1, 8-2), respectively, from each PCR were sequenced partially and turned out to be identical. The corresponding locus was named Pptub 1.

This preliminary sequence information was used to design primers

in order to perform a genomic walk into flanking regions of Pptub 1, using an I-PCR approach on religated EcoR I and Hind III genomic digests (primers 35, 36). Reconfirmation of products was done by nested PCR (primers 40, 38). Two clones generated by nested PCR products (E#1 and H 1.7) were sequenced completely.

The Hind III clone H 1.7 did not harbour an internal Hind III site, most likely due to star activity of the enzyme or ligation of a random ds breakage. However, sequences upstream of the first EcoR I site were confirmed by two independent PCRs on genomic DNA (primers 113, 67 and 113, 90). In addition, an additional cDNA (89, 91; Pfu native (MBI Fermentas, St. Leon-Rot, Germany)) PCR product was cloned.

All mentioned clones helped to generate and reconfirm sequence data. In total ~1500 bp upstream of the startcodon and ~1500 bp downstream of the stopcodon were gained.

Pptub 2

As already described above sequence information of published ESTs from Physcomitrella patens was used to design a gene-specific primer (F7) upstream of the predicted coding region. PCR on Physcomitrella patens genomic DNA (primers F7, 10R) and subsequent cloning and sequencing of the PCR product proofed that it, together with all three so far published Pptub ESTs (Pptub EST 1-3) belong to one locus, named Pptub 2. Intron positions could be verified by comparing EST with genomic sequences.

This preliminary sequence information was used to design gene-specific primers within introns (primers 95 and 71) in order to perform a genomic walk into adjacent genomic regions of Pptub 2, using an I-PCR approach on religated Pag I, BamH I and Nde I genomic digests. PCR products were reconfirmed by nested PCR (primers 38, 35). Two clones generated by nested PCR products (C#2Pag and D#2Nde) were sequenced completely. The Nde I clone D#2 did not harbour an internal Nde I site, most likely due to star activity of the enzyme or ligation of a random ds breakage. However, sequence data were confirmed by C#2Pag and a third I-PCR clone (95#8BamHI; primer 149 and 71). In addition two independent PCRs on genomic DNA (primers 205, 149; Taq (MBI Fer-

mentas, St. Leon-Rot, Germany) and primers 205, 206) confirmed product length. The 205-206 PCR product and an additional genomic downstream PCR product (primers 71, 206; Pfu native (MBI Fermentas, St. Leon-Rot, Germany)) were cloned and helped to verify sequence data.

All mentioned clones helped to generate and reconfirm sequence data. In total ~1400 bp upstream of the startcodon and ~1400 bp downstream of the stopcodon were gained.

Pptub 3

As already mentioned in a first step, Taq (MBI Fermentas, St. Leon-Rot, Germany) PCR fragments from two independent PCRs on Physcomitrella patens genomic and cDNA using primers 9F and 10R were cloned. Clones from each PCR (#3-3 genomic, #4-3 cDNA) were sequenced partially and turned out to be identical. The corresponding locus was named Pptub 3.

This preliminary sequence information was used to design gene-specific primers within introns (primers 69, 70) in order to perform a genomic walk into adjacent regions of Pptub 3, using an I-PCR approach on religated Pag I and Nco I genomic digests. Reconfirmation of PCR products was done by nested PCR (primers 38, 35). Two clones (A#1Nco and #4-1Pag) were sequenced completely. A#1Nco is a clone generated by a nested PCR product (38, 35) whereas #4-1PagI was generated by the original I-PCR product (69, 70). In addition a genomic PCR product (primers 203, 204) was cloned and helped to verify sequence data.

All mentioned clones helped to generate and reconfirm sequence data. In total ~1900 bp upstream of the startcodon and ~1100 bp downstream of the stopcodon were gained.

Pptub 4

As already mentioned, in case of Pptub 4, EST data were used to design gene-specific downstream and upstream primers (297, 299) in order to generate genomic and cDNA clones. Additional genomic clones using native Pfu polymerase (MBI Fermentas, St. Leon-Rot, Germany) helped to verify sequence data.

Primer 297 and 299 were inverted (337, 383) and used to perform a walk into adjacent genomic regions of Pptub 4, using an I-PCR approach on religated Nde I and Nco I genomic digests. Two clones (48#2Nco and A02#3Nde) and additional genomic clones (primers 547 and 374; Advantage cDNA Polymerase Mix (BD Biosciences Clontech, Heidelberg, Germany) and Triple Master (Eppendorf, Hamburg, Germany)) were generated.

All mentioned clones helped to generate and reconfirm sequence data. In total ~2300 bp upstream of the startcodon and ~1100 bp downstream of the stopcodon were gained.

Pptub 5 and 6

As already mentioned, in case of Pptub 5 and 6, EST data were used to design gene-specific downstream and upstream primers (Pptub 5: 298, 300 and Pptub 6: 296, 336) in order to generate genomic and cDNA clones of each gene. In case of Pptub 5, additional genomic clones using native Pfu polymerase (MBI Fermentas, St. Leon-Rot, Germany) helped to verify sequence data.

All mentioned clones helped to generate and reconfirm sequence data. In total 2031 bp genomic sequence for Pptub 5 and 3161 bp genomic sequence for Pptub 6 were gained.

Cloning strategies

Preliminary Pptub 1 (2-1, 8-1, 8-1; all genomic) and Pptub 3 (3-3 genomic, 4-3 cDNA) clones were generated with Taq recombinant polymerase. PCR products were ligated into TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany). PCR conditions were: 2.5 unit Taq recombinant polymerase, enzyme buffer, 3.3 mM $MgCl_2$ (all MBI Fermentas, St. Leon-Rot, Germany), 0.4 mM each primer, 100 nanograms of cDNA or genomic DNA as template in a total volume of 25 microliters. Cycling conditions were: an initial step of 5 minutes at 95°C, then 45 seconds 95°C, 10 seconds 60°C (primer 8F) or 65°C (primer 9F) and 1 minute 72°C as a second step, with 30 to 35 repetitions, followed by a terminal step of 5 minutes at 72°C and cooling to 4°C at the end of the program.

All other genomic and cDNA clones were

Pptub 1: 113-67, 113-90, 89-90, 89-91 cDNA
Pptub 2: F7/R10, 205-206, 71-206
Pptub 3: 203-204
Pptub 4: 547-374 (+ TrippleMaster), 297-299 cDNA + genomic (+ Pfu)
Pptub 5: 298-300 cDNA + genomic (+ Pfu)
Pptub 6: 296-336 cDNA + genomic

Underlined clones above were generated with Advantage cDNA Polymerase Mix, using 0.25 microliters enzyme mix, buffer (including 3,5 mM $Mg(OAc)_2$, both BD Biosciences Clontech, Heidelberg, Germany), 0.25 mM each primer, 0.25 mM dNTPs and 10-20 nanograms of template per 20 microliter PCR. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96°C, 10 seconds 60°C and 2 minutes/kb 68°C as a second step, with 35 to 40 repetitions, followed by a terminal step of 15 minutes at 68°C and cooling to 4°C at the end of the program. PCR products of appropiate length were eluted from agarose gels. Elution was done in 30-50 microliters, depending on amount of amplificate. Eluted PCR products were cloned in TOPO TA vectors (pCR4-TOPO,Invitrogen, Karlsruhe, Germany).

All other clones were generated with Pfu native polymerase, as were the two additional genomic clones 297-299 and 298-300, using 0.3 microliters polymerase (= 0.75 units), buffer, 2-4 mM $MgSO_4$ (all MBI Fermentas, St. Leon-Rot, Germany), 0.25 mM each primer, 0.2 mM dNTPs and 10-20 nanograms of template per 20 microliter PCR. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96°C, 10 seconds 60°C and 2 minutes/kb 72°C as a second step, with 35 to 40 repetitions, followed by a terminal step of 10 minutes at 72°C and cooling to 4°C at the end of the program. PCR products of appropriate length were eluted from agarose gels. Elution was done in 30-50 microliters, depending on amount of amplificate. Eluted PCR products were cloned in pZErO-2 (Invitrogen, Karlsruhe, Germany) linearised with EcoRV.

An additional clone of 547-374 was generated with the TripleMaster PCR System, using 0.25 microliters polymerase mix (= 1.25

units), tuning buffer (including 2.5 mM Mg^{2+}, both Eppendorf, Hamburg, Germany), 0.2 mM each primer, 0.2 mM dNTPs and 10-20 nanograms of template per 20 microliter PCR. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96° C, 20 seconds 60°C and 3 minutes 72°C as a second step, with 40 repetitions, followed by a terminal step of 10 minutes at 72°C and cooling to 4°C at the end of the program. PCR products of appropriate length were eluted from agarose gels. Elution was done in 30-50 microliters, depending on amount of amplificate. Eluted PCR products were cloned in TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany).

In summary, PCR on genomic DNA of Physcomitrella patens and cloning of PCR products lead to sequence information of six transcribed Physcomitrella patens ß-tubulin genes. Additionally, EST and cDNA data were used to confirm genomic sequence data and intron/exon borders. In case of Pptub 1 to 4 inverse PCR lead to non transcribed flanking 5´ and 3´ genomic sequences. A general overview of all six genomic regions is given in Figure 1.

Gene structure & Conservation

As already stressed, Pptub 1 to 4 are most abundantly represented in EST databases. In addition the great majority of their corresponding ESTs were raised from full length cDNA libraries. This two facts helped to determine the transcriptional start site (TSS) of Pptub 1 to 4 *in silico*. A multiple alignment of 5´ ESTs against corresponding upstream genomic regions showed that Pptub 1to 3 do have a precise transcriptional initiation: 20 out of 27 5´ ESTs for Pptub 1, 16 out of 20 5´ ESTs for Pptub 2 and 9 out of 14 5´ ESTs for Pptub 3, do start at the same, most upstream position, marked with +1 (Figure 3-6). In addition all three TSSs are surrounded by a consensus sequence (see below). In case of Pptub 4 the 23 5´ ESTs indicate multiple TSSs within 100 bp. The start site of the most upstream 5´ EST was defined as +1.

An analogous multiple alignment of 3´ ESTs against corresponding downstream genomic regions reconfirmed that plant genes almost always come with more than one poly(A) site and that consensus

sequences are much less sharply defined than in e.g. mammalian genes, in which the sequence AAUAAA is nearly ubiquitous (for review see: Rothnie et al., 1996).

The six cloned loci of Physcomitrella patens did not show any nonsense stop-codons and proper proteins with high similarities to known ß-tubulins could be predicted. Outside the coding regions generally, the similarity drops immediately and significantly. Concerning 5´ putative regulatory elements, a detailed comparison of all four upstream regions revealed no overall conservation within the gene family or to 5´ regions of other known plant ß-tubulin genes. However, some interesting matches of conservation within the gene family could be detected:

a) The determined TSSs of Pptub 1 to 3 in all three cases fall within the consensus sequence T/C C A(+1) G/C T G T G C and are embedded in C/T-rich regions (compare consensus of 171 unrelated TATA plant promoters: T/C C A(+1) N M N in plantProm Database under http://mendel.cs.rhul.ac.uk/mendel.php?topic=plantprom).
b) 22-24 bp upstream of the TSS -which is within the typical distance for plant TATA promoters (see plantProm DB)- a weak 8 bp TATA box embedded in a conserved stretch of 20-25 bp can be found in Pptub 1 to 3 . The TATA box consensus from 171 unrelated plant promoters is: T_{96} A_{95} T_{96} A_{100} A_{62}/T_{38} A_{97} T_{61}/A_{38} A_{73} (see plantProm DB) and for Pptub 1-3 is: T t T A T c T c/t/A, with capitals indicating correlation to consensus.
c) all four genes do have a very low degree of Adenosine (9-16%) in their 5´UTRs.
d) The 5´ UTR of Pptub 4 has an overall C/T content of 74%, which -in addition- harbours a C/T stretch (~ 50 bp), directly behind the start point of the shortest, most downstream 5´ EST.
e) Pptub 2 harbours a 40 bp polyA stretch around 450 bp upstream of the TSS (-450 until -489).
f) In Pptub 1 and 4 upstream of app. position -420 long very A/T-rich regions begin (Pptub 1 over 80% A/T for nearly 900 bp and Pptub 4 75% A/T for 1750 bp , rendering open the possibility for the location scaffold/matrix attached regions (S/MARs; (Liebich et al., 2002) upstream of this genes.

Functional Characterization & Quantification of ß-tubulin pro-

moters

Definition of minimal promoter-fragments giving a maximum of promoter activity was done by functional quantification of putative 5´ regulatory sequences of Pptub 1 to 4 in a transient expression system, using nonregenerating Physcomitrella patens protoplasts as expression system. For each promoter several constructs of different lengths including upstream regions and 5´ UTRs, were brought precisely in front of the startcodon of the reporter gene. As reporter gene a human protein (recombinant human vascular endothelial growth factor 121: rhVEGF121; Gorr 1999) was secreted into the medium via its own signalpeptide. The amount of rhVEGF121 in the supernatant of the moss culture was quantified by an ELISA and reflected the strength of the promoter or promoter fragment in the system. Values were related to values obtained by the 35S promoter. Each construct was transformed a minimum of six times in two to three different transformation experiments. Samples were taken after 24 and 48 hours, respectively, with 48 hour samples measured twice in appropriate dilutions in an ELISA. An overview of the results is given in Figure 2 .

The expression promoting regions of Pptub 1 to 4 are disclosed as Seq.ID.Nos. 1 to 8.

Cloning of amplified promoter fragments of Pptub 1 and 4 into pRT101new

Pptub 1: 1-0 (primer 364XhoI, 363cat)
1-1 (primer 219XhoI, 363cat)
1-3 (primer 549XhoI, 363cat)
1-4 (primer 226XhoI, 363cat)
1-5 (primer 550XhoI, 363cat)

Pptub 2: 2-0 (primer 291, 225cat)

Pptub 3: 3-0 (primer 292, 223cat)

Pptub 4: 4-0 (primer 373XhoI, 374cat)
4-1 (primer 548XhoI, 374cat)

The promoter fragments given above were amplified with Pfu native polymerase (MBI Fermentas, St. Leon-Rot, Germany) on genomic DNA using reverse primers starting with the reverse complement sequence of the ATG start codon (cat...) and, in part, forward primers containing XhoI sites. PCR products were cut XhoI and ligated into XhoI/HincII or not cut at all and and ligated into HincII opened pRT101new, respectively. Generated clones were verified by sequencing. Clone 1-2 (XhoI/EcoRI), 2-1 (BglII), 2-2 (SalI), 2-3 (EcoRI/SalI), 2-4 (EcoRI/SalI), 3-2 (SalI), 3-3 (Eco147I/HincII), 3-4 (XhoI/SalI) were generated by internal deletions of longer clones. The remaining vectors were gel-eluted and religated. In case single strand overhangs did not fit, ligation was performed after filling-in of recessed 3´-termini with Klenow Fragment (MBI Fermentas, St. Leon-Rot, Germany), following the suppliers manual.

Pptub 1

Six different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter gene. The data of all constructs are given in figure 3. (5´ UTR = +1 (TSS) until +226, +227= ATG)

1-0	-1307 bp	(1533 bp 5´ region of Pptub 1)
1-1	- 985 bp	(1211 bp 5´ region of Pptub 1)
1-2	- 416 bp	(642 bp 5´ region of Pptub 1)
1-3	- 248 bp	(474 bp 5´ region of Pptub 1)
1-4	- 83 bp	(309 bp 5´ region of Pptub 1)
1-5	- 71 bp	(297 bp 5´ region of Pptub 1)

Promoter fragment 1-2 can be defined as the shortest promoter fragment giving high expression rates. The rates are app. 150% compared to values generated with the 35S promoter, which was set to 100%. Note that upstream of the minimal promoter fragment 1-2 a long, very A/T rich region starts (over 80% A/T for nearly 900 bp).

Pptub 2

Five different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter

gene. The data of all constructs are given in Figure 4. (5´ UTR = +1 (TSS) until +122, +123= ATG)

2-0	-1075 bp	(1197 bp 5´ region of Pptub 2)
2-1	- 676 bp	(798 bp 5´ region of Pptub 2)
2-2	- 425 bp	(547 bp 5´ region of Pptub 2)
2-3	- 245 bp	(367 bp 5´ region of Pptub 2)
2-4	- 67 bp	(189 bp 5´ region of Pptub 2)

Promoter fragment 2-2 can be defined as the shortest promoter fragment giving high expression rates. The rates are comparable to values generated with the 35S promoter (100%).

Pptub 3

Different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter gene. The data of four constructs are given in Figure 5. (5´ UTR = +1 (TSS) until +112, +113= ATG)

3-0	-1274 bp	(1386 bp 5´ region of Pptub 3)
3-2	- 765 bp	(879 bp 5´ region of Pptub 3)
3-3	- 272 bp	(384 bp 5´ region of Pptub 3)
3-4	+ 52 bp	(60 bp 5´ UTR of Pptub 3)

Promoter fragment 3-2 can be defined as the shortest promoter fragment giving high expression rates. The rates are app. 300% compared to values generated with the 35S promoter, which was set to 100%.

Pptub 4

Two different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter gene. The data are given in Figure 6.
(5´ UTR = TSSs (+1 until +103) until +205, +206= ATG)

4-0	-419 bp	(624 bp 5´ region of Pptub 4)
4-1	- 1 bp	(206 bp 5´ region of Pptub 4)

Promoter fragment 4-1 gives expression rates that are are app. 250% compared to values generated with the 35S promoter, which

was set to 100%. Note that upstream of this minimal promoter fragment (4-0) a long, very A/T rich region starts (75% A/T for 1750 bp).

In summary transient promoter activity of Pptub 1 to 4 genomic upstream regions were characterised. Minimal promoter fragments showing a maximum of promoter activity were defined and gave yields of up to 3 times the 35S promoter activity.

Pptub-constructs summary (see also: sequence listing)

Pptub1 upstream
-1533 until -1 (+1 = start codon)
-1533 until -644 = 81 % AT
-1533 VEGF 1-0 (primer 364)
-1211 VEGF 1-1 (primer 219)
-642 VEGF 1-2 (EcoRI/XhoI)
-474 VEGF 1-3 (primer 549)
-309 VEGF 1-4 (primer 226)
-297 VEGF 1-5 (primer 550; without putative TATA box: -304 until -295)
-226 TSS (start of 5´UTR)

Pptub1 downstream
1 until 1539 (1 = directly behind stop codon)
332 end of longest EST (3´UTR)
1539 start of primer 90

Pptub2 upstream
-1197 until -1 (+1 = start codon)
-1197 VEGF 2-0 (primer 291)
-798 VEGF 2-1 (BglII)
-547 VEGF 2-2 (SalI)
-450 until -489 = poly A stretch
-367 VEGF 2-3 (EcoRI/SalI)
-189 VEGF 2-4 (XhoI/SalI)
-122 TSS (start of 5´UTR)

Pptub2 downstream
1 until 1012 (1 = directly behind stop codon)

297 end of longest EST (3´UTR)
1012 start of primer 206

Pptub3 upstream
-1386 until -1 (+1 = start codon)
-1386 VEGF 3-0 (primer 292)
-879 VEGF 3-2 (SalI)
-384 VEGF 3-3 (Eco147I/HincII)
-112 TSS (start of 5´UTR)
-60 VEGF 2-4 (XhoI/SalI)

Pptub3 downstream
1 until 997 (1 = directly behind stop codon)
203 end of longest EST (3´UTR)
1012 start of primer 204

Pptub4 upstream
-624 until -1 (+1 = start codon)
-624 VEGF 4-1 (primer 373)
-206 VEGF 4-2 (primer 548)
-205 until -103 area of TSS (start of 5´UTR)
-55 until -93 CT stretch

Pptub4 downstream
1 until 1146 (1 = directly behind stop codon)
466 end of longest EST (3´UTR)
1141 until 1164 NcoI

EXAMPLE 2: Cloning and analysis of actin genes from different moss species and their expression promoting regions.

2.1.Genomic structure of Physcomitrella patens actin genes.

Four actin genes and promoter regions of the moss Physcomitrella patens and three from Funaria hygrometrica and the liverwort Marchantia polymorpha have been isolated in order to construct expression vectors for their use in moss.

Using specific oligos designed from Physcomitrella EST sequences that are present in the public databases, four actin genes

(Ppact1, Ppact3, Ppact5 and Ppact7) were isolated in several rounds of iPCR from genomic DNA and sequenced.

In Physcomitrella the structure of the isolated genes resembles in one case (Ppact1) the conserved structural organisation of actin genes of higher plants. The un-translated leader is disrupted by a relatively long (955bp) intron located 14 nt upstream the initiator ATG. The coding region presents three smaller introns which are situated at the same positions as the introns of actin genes of other plant species. The first one is located between codons 20 (lys) and 21 (ala), the second is splitting codon 152 (gly) and the third is between codon 356 (gln) and 357 (met). This general structure appear to be different for the three other Physcomitrella actin genes isolated (Ppact3, Ppact5, and Ppact7). In those cases the 5´UTR intron (434bp, 1006bp and 1055bp respectively) is also located 14nt before the ATG but the coding region is disrupted only by one intron positioned between codons 21 (lys) and 22 (ala) (Fig. 7).

2.2.Activity studies of the expression promoting regions of actin genes.

To study the activity of the different Physcomitrella actin expression promoting regions (Seq.ID Nos. 5 to 8) as well as the effect of the 5'UTR of the different genes, different vectors were designed for expression of the hVEGF protein under the control of the 5´ regions under study.

Around 2kb genomic regions upstream the transcription initiation site were isolated by iPCR from genomic DNA and sequenced, and vectors containing the cDNA of the human VEGF driven by the promoters and containing the exact leader sequences including the 5´intron were constructed for transient transfection of moss protoplasts. The complete 5´promoting expression regions were amplified by proof reading PCR using primer 395 and 332 for Ppact1, 408 and 333 for Ppact3, 511 and 334 for Ppact5, and 413 and 335 for Ppact7.

Transformation of protoplasts was performed using the same number of molecules for each construct to be tested and in parallel

to a construct carrying the hVEGF cDNA under the control of the CaMV 35S promoter. The hVEGF protein contains at the N-terminal part a 26 aa signal peptide that permits secretion of the recombinant protein to the medium. Analysis of the transformations was carried out by ELISA, taking different dilutions of the medium where the protoplasts were incubated 48 hours after transformation.

The capacity to drive expression of the different Physcomitrella 5´ actin regions was compared to the activity of the constitutive 35S promoter.

In all cases analysed, the 5´regions of the actin genes were reaching higher activity than the 35S promoter. However the level of expression varied for the different actin regulatory sequences. Thus, the 5´sequence of Ppact3 was only promoting around a 2 fold higher expression of VEGF than the 35S promoter. Higher levels of VEGF were measured when vectors containing the 5´regions of Ppact1 and Ppact7 were used for transformation. In those cases values between 4 and 8 folds the 35S values were obtained. Nevertheless the most dramatic differences were observed in the case of the 5´Ppact5 gene, where up to 11 fold higher expression values compared to the 35S were in some cases obtained (Fig. 8).

To further investigate on the role of the 5´ UTR region of the high activity Physcomitrella actin genes, vectors containing deletions, combinations and substitutions of the 5´UTR intron were made and used for transient assays in moss protoplasts.

Deletion of the Ppact1 5´intron dramatically decreased the levels of transient expression in comparison to those obtained when the intact 5´region of Ppact1 was used. In this case the amount of secreted VEGF protein that could be detected in the protoplasts medium was very similar to the obtained by the CaMV 35S promoter. This would indicate that the 5´intron of the Ppact1 is essential for efficient gene expression from the Ppact1 promoter. Same results were obtained when the 5´UTR including the leader intron was fused downstream the 35S promoter. This construct yielded the same amount of secreted protein as

the intact 35S promoter indicating that the 5'UTR region is not having any dramatic influence on the activity of promoters other than the Ppact1 promoter. It is important to indicate that a construct carrying just the 5'UTR Ppact1 region was able to promote protein production only in a 30% lower amount than the 35S promoter alone. This could suggest a small promoter activity in this region of the gene, or a rest of promoter activity present in the backbone sequence of the vector (Fig. 9).

The same approach was used to investigate the influence on the promoter activities of the 5'UTR introns contained in the Ppact5 and Ppact7 genes. Constructs in which the 5' intron was deleted were analysed and similar results as in the case of Ppact1 were obtained, ie. the amount of protein reached was approximately the same as with the 35S promoter in the case of Ppact5 and slightly lower in the case of Ppact7, indicating that the presence of the intron in the 5'UTR is essential for the efficient activity of the promoters. Again some residual promoting activity was observed when the transformation was performed with constructs containing only the 5'transcribed region up to the ATG. Furthermore, in the case of these two genes, the fusion of the 5 'UTR downstream the 35S promoter yielded higher rates (2 to 7 folds) of expression of the VEGF protein when compared to the 35S promoter alone (Fig. 10, 11). Similar results were observed in the case of Ppact3, where the 5'UTR alone or fused downstream the CaMV 35S, yielded around 2 and 3 folds respectively in comparison to the 35S (Fig. 12). These indications would suggest the presence of enhancer activity in the 5'transcribed regions for these three genes even when they are positioned under a different promoter.

To further investigate the role of the 5'intron present in the Ppact1, Ppact5 and Ppact7 genes, substitutions of the leader intron of the Ppact1 gene with the 5'intron of Ppact5 and Ppact7 were engineered in vectors for transient transformation. In parallel substitutions of the Ppact1 5'intron with the ppact1 introns present in the coding region of the gene, were performed.

Substitutions of the Ppact1 5'intron, by the Ppact 1 coding region introns 1 and 3 resulted in a decrease of the expression

levels of around 25%. Still the amount of protein detected was around 2-3 fold higher than the obtained with the CaMV 35S promoter. The substitution of the 5´intron by the intron 2 of the coding region surprisingly resulted in no activity of the promoter (Fig. 13). The construct was however checked, and the sequence showed that the splicing site for the intron was not correct. A new construct carrying the correct splicing sequence was made and the results after moss transformation indicated that the effect of the intron 2 is the same as for the other substitutions.

A reduction of protein expression was also observed when the substitution was done with the 5´introns corresponding to the Ppact5 and Ppact7 genes, but in this case the reduction was slightly smaller.

2.3. Deletion constructs of the expression promoting regions of actin genes.

A further characterisation of the different actin genes promoters was carried out by making deletion constructs of the 5´untranscribed regions and analysing them through transient transformation of moss protoplasts.

Thus for the Ppact1 constructs carrying different genomic region lengths (-1823bp, -992bp, -790bp, -569bp, -383bp, -237bp, and -82bp) upstream the initiation of transcription (+1) were made. In principle all the constructs except the -82bp, could have full promoter activity. However the -383bp construct shows a reduction of activity and reaches similar levels as the -82bp construct (Fig.14).

Analysis of deletion constructs of the promoter region of Ppact3 revealed some interesting features. As it was described, this promoter presented a lower activity compared to the other actin genes promoters, although in relation to the CaMV 35S, it was slightly more active. In this case the following 5´untranscribed regions were tested: -2210bp, -995bp, -821bp, -523bp, -323bp, -182bp and -81bp. Surprisingly the activity of the promoter was approximately the same as the CaMV 35S for the constructs con-

taining up to -821bp of the promoter region. However the constructs containing from bp -523 and shorter regions towards the transcription start, yielded two folds more amount of recombinant protein. This could indicate cis-acting regions located upstream the -523bp region that down regulate the transcription of this gene during the transient transformation assay (Fig. 15).

In the case of Ppact5, constructs containing the -1872bp, -758bp, -544bp, -355bp, and -121bp fragments upstream the transcription start of the gene were generated. The results obtained from the transient assays indicate that the full activity of the promoter resides in a region between -758 and -121 from the start of transcription (+1) (Fig. 16).

The following deletion constructs for the 5´untranscribed region of Ppact7 were analysed: -1790bp, -1070bp, -854bp, -659bp, -484bp, -299bp, and -66bp. The results obtained indicate that the region comprised in between -484bp and -299bp is essential for the full activity of the promoter during the transient experiment assays. (Fig.17).

In order to obtain a set of heterologous promoters of the Physcomitrella actin genes, other two species, the moss Funaria hygrometrica and the liverwort Marchantia polymorpha, were used to isolate genomic DNA fragments containing actin genes. To this end, oligos with different degrees of degeneration were designed to perform PCR reactions using as template genomic DNA isolated from the two species.

2.4.Comparison of different actin genes from the different moss species Physcomitrella patens, Funaria hygrometrica and Marchantia polymorpha

Physcomitrella patens

The four different genomic actin sequences isolated from Physcomitrella patens are likely to represent the whole functional sequences of the genes including 5´promoter sequence, 5´UTR + 5´intron, ORF + internal introns and the 3´UTR and further 3´downstream sequence. In total for Ppact1 5809 bp, for Ppact3

5633 bp, for Ppact5 8653 bp and for Ppact7 6351 bp of genomic sequence was isolated (Fig. 18 A). The coding regions of the isolated Physcomitrella actin cDNAs are almost all 1137 bp in length, except Ppact1 which has an ORF of 1134 bp. The corresponding proteins are 378 amino acids in lengths except Ppact1 which has 377 amino acids. On the nucleotide level the coding sequences share homologies between 86.6 and 98.9 %. The protein sequences have an identity between 97.1 and 99.7 % (DNA STAR, MegAlign Program, Clustal V (weighted) sequence alignment).

For all four Physcomitrella actin genes extended genomic DNA sequences 5´of the ATG Start codon could be isolated by iPCR and sequenced: 2973 nt for Ppact1, 3091 nt for Ppact3, 3095 nt for Ppact5 and 3069 nt for Ppact7. For Ppact1, Ppact5 and Ppact7 5´race by using the Gene Racer Kit (Invitrogen), which allows the amplification of only full length cDNAs, was performed to determine the 5´UTRs of the genes. For Ppact3 the 5´UTR was determined by the length of different ESTs from database. By comparing the cDNAs with the genomic iPCR fragments the presence of large 5´introns could be shown. The lengths of the 5´introns which are all located at position -14 to the ATG Start codon are 955 bp, 434 bp, 1006 bp and 1055 bp for Ppact1, Ppact3, Ppact5 and Ppact7 respectively (Fig. 18 A). The positions of the ORF internal introns was determined by comparing the genomic sequences and the derived protein sequences to the cDNA sequences and protein sequences of the actin genes from Arabidopsis thaliana. The 5´promoter sequences for the Physcomitrella actin genes available are 1824 nt for Ppact1, 2270 nt for Ppact3, 1909 bp for Ppact5 and 1805 bp for Ppact7 (Fig. 18 A).

In total 4 different actin genes from Funaria hygrometrica (expression promoting regions: Seq.ID Nos. 9 to 12) and 3 different genes from Marchantia polymorpha (expression promoting regions: Seq.ID Nos. 13 to 15) could be identified by degenerated PCR on genomic DNA. As the aim was predominantly to isolate 5´promoter regions of the putative different actin gene homologs from the different moss species, most of the sequences are incomplete at the 3´end to date (Fig. 18 B/C).

Funaria hygrometrica

For Funaria the identified actin genes were named Fhact1, Fhact4.4, Fhact5 and Fhact5b. 3951 bp of Fhact1, 2417 bp of Fhact4.4, 4432 bp of Fhact5 and 722 bp of Fhact5b of genomic sequence could be isolated by iPCR for the different actin genes. The complete coding cDNA sequence could be isolated for the Fhact1 gene which has a coding sequence of 1134 nucleotides. For the other Funaria actin genes partial sequences are available at the moment, lacking the 3´ends: 906 bp for Fhact4.4, 965 bp for Fhact5 and 722 bp for Fhact5b (Fig. 18 B) The isolated coding sequences share homologies in a range of 87.4 and 99.2% on the nucleotide level. The derived protein sequences are 90.8 to 99.2 % identical (DNA STAR, MegAlign Program, Clustal V (weighted) sequence alignment).

Except for Fhact5b, 5´sequences upstream of the ATG Start codon could be isolated by iPCR and sequenced. In the case of Fhact1 1824 bp, for Fhact4.4 1333 bp and for Fhact5 3289 bp are available. The length of the different 5´UTRs were determined by 5 ´race using the Gene Racer Kit (Invitrogen). The intron-exon structure was determined by comparison of the cDNA sequence with the genomic sequences obtained by iPCR and by comparison to the Physcomitrella genes. As in the case of the Physcomitrella actin genes the identified Funaria actin genes contain large 5 ´introns located at position -14 of the cDNAs, 928 bp, 1015 bp and 656 bp in length for Fhact1, Fhact4.4 and Fhact5 respectively. By now for Fhact1 700 bp, 145 bp for Fhact4.4 and for Fhact5 2515 bp of 5´promoter sequence was isolated and sequenced. For Fhact1 419 bp of the 3'region was isolated. The 5'regions or 3'regions of the Funaria actin genes are amplified by PCR on genomic DNA from Funaria hygrometrica by using the primers 908 and 909 for the 5' region of Fhact1, 983 and 984 for the 3' region of Fhact1, 1000 and 1001 for the 5'region of Fhact4.4 and 611 and 612 for the 5'region of Fhact5.

Marchantia polymorpha

For Marchantia the identified actin genes were named Mpact1, Mpact4 and Mpact15. For all three sequences the 3´ends are lacking. So far for Mpact1 2229 bp, for Mpact4 3987 bp and for

Mpact15 2174 bp of genomic sequences were isolated and sequenced. The lengths of the coding cDNA sequences isolated are 997 nt, 962 nt and 995 nt for Mpact1, Mpact4 and Mpact15 respectively. (Fig. 18 C). The sequence homologies within the Marchantia actin genes are a little bit lower than compared to the other two moss species, in a range between 78.3 and 85.5 % on the nucleotide level and between 94.7 and 96.1 % on the amino acid level (DNA STAR, MegAlign Program, Clustal V (weighted) sequence alignment). 5´upsteam sequence of the ATG for all the three identified different Marchantia actin genes were isolated by iPCR and sequenced: 937 bp for Mpact1, 3025 bp for Mpact4 and 910 bp for Mpact15. The 5'regions of the the Marchantia actin gene homologous are amplified by PCR on genomic DNA from Marchantia polymorpha using the primer 950 and 951 for 5'Mpact1, 960 and 961 for Mpact4 and 970 and 971 for Mpact15. The intron-exon structure of the ORF was obtained by comparing the different actin gene sequences from the different moss species. The isolated 5´ sequence of Mpact1 shows the consensus sequence for intron splice sites (aggt) at position -14 indicating the presence of a 5´intron as in the case of the other Physcomitrella and Funaria genes. Within the 5´upstream sequences of Mpact4 and Mpact15 no intron splice site consensus sequence is present, proposing the lack of 5´introns (fig. 18 C).

Comparison of of P. patens, F. hygrometrica and M. polymorpha actin genes

As mentioned above in general the homologies of nucleotide and protein sequences for the different isolated actin genes within one species is very high especially at the protein level. The homologies between the closely related moss species Physcomitrella patens and Funaria hygrometrica also appear to be very high. On the nucleotide level the actin genes show homologies between 86.9 and 96.3 % identity and on the amino acid level the range of homology is 95.5 to 99.7 %.

In contrast to that the more distant relation of the liverwort Marchantia polymorpha to the other both species is reflected in the lower homologies of the genes on the nucleotide level. The homologies between Physcomitrella and Marchantia actin genes is

in the range of only 75.2 % and 78.8 % and between Funaria and Marchantia the homologies are in the range of 75.5 % to 80.4 %. On the amino acid level the homologies of the Marchantia actin genes vary between 93.0 % and 96.1 % compared to Physcomitrella and between 93.4 % and 96.7 % compared to Funaria.

Intron-exon structure (Fig. 18 A/B/C)

As indicated before the intron-exon structure of the Physcomitrella actin genes to a certain extent are similar to that of higher plants but also with clear differences. All isolated Physcomitrella actin genes contain a large 5´intron in the 5´untranslated region, which almost all of the investigated higher plants actins do. Only Ppact1 contains 3 internal introns within the ORF reflecting the situation for example for all isolated actin genes from Arabidopsis thaliana. The ORF internal intron positions of Ppact1 are also conserved compared to higher plant actin genes. On the contrary Ppact3, Ppact5 and Ppact7 contain only one internal intron within the ORF.

The same genomic structure can be found in the isolated Funaria actin genes with one extended 5´intron within the 5´UTR. Fhact1 has the same conserved intron-exon structure as Ppact1 whereas Fhact4.4 and Fhact5 contain only one internal intron within the ORF sequence. The isolated sequence of Fhact5b is to short to say something clear about the intron-exon structure but at least it does not contain the internal intron2 compared to Fhact1 or Ppact1.

In Marchantia the genomic structures of the isolated actin genes seem to be more different. It is important though, to indicate that the number of different actin genes in the three different moss species is not known and it could be that the three isolated actin genes from Marchantia do not represent the individual functionally homologous genes. It is likely that there are more than three actin genes present in Marchantia and more than four actin genes in Physcomitrella and Funaria.

However, the intron-exon structure of Mpact1 seems to be the same as in the case of Ppact1 and Fhact1 with a 5´intron within

the 5´UTR and the conserved positions of the ORF internal introns 1 and 2. Mpact15 also contains the conserved ORF internal intron1 and intron2 but it does not have a conserved intron splice site at position -14 within the 5´UTR or at position -10 as found for the Physcomitrella or for some Arabidopsis actin 5 ´introns respectively, arguing for a lack of a 5´intron. The same situation is found for Mpact4, probably lacking a 5´intron. In addition Mpact4 also does not have the intron1 or the intron2 within the ORF, which is different from all isolated moss actin genes so far.

Putative homologous moss actin genes

Although the intron-exon structure of the different isolated actin genes from Physcomitrella and Funaria might propose conclusions about homologous genes between the two species one can not conclude this from the genomic structure. For example Ppact1 and Fhact1 share the same conserved intron-exon structure but it is not clear, as indicated before, whether there are more genes present in the genome of both plants which might have the same genomic structures. To give a statement on homologous genes also expression data would be required to propose functional homologies. Also from the sequence homologies of the proteins or the coding cDNA sequences it is not possible to make any assumptions about corresponding homologous genes between the species as they are too similar in general.

But in the case of Physcomitrella and Funaria it was interesting to find also very high sequence homologies within the non coding sequences regarding to the UTR sequences, intron sequences and promoter sequences. Therefore high homologies were found between Ppact1 and Fhact1 and between Ppact3 and Fhact5. In both cases the intron sequences showed unusual high conservation. In the case of Ppact1 and Fhact1 the homologies were as follows: 5´intron: 58 %; intron1: 64 %, intron2: 52 % and intron3: 55 %. In the case of Ppact3 and Fhact5 the homologies are for the 5´intron 51 % and intron1 shows 48 % identity.

For both cases also the isolated 5´ promoter sequences show high homologies. Fig. 19 A shows a schematic comparison of the isol-

ated promoter regions of Ppact1 and Fhact1. The transcription start is said to be at position 1, the first nt of the 5´promoter region is said to be -1. The isolated 267 bp of 5´promoter region of Fhact1 show an over all homology to the first 267 bp of the Ppact1 5´promoter region of 58 %. Within this sequence there are blocks of different homologies observable. The sequence between -267 and -129 shows a homology of 51 %. The following 29 bp show 62 % identity and within position -100 and -1 the homology is almost 70 %. Concerning these high sequence identities between the Ppact1 and Fhact1 intron and promoter sequences it is reasonable to put these two genes as the homologous genes in these two mosses. Another interesting aspect is the observation of the drop of expression observed between the different Ppact1:vegf deletion constructs (Fig. 15). The dramatic drop of expression appears to be between the -237 and the -82 deletion construct. This argues for an important function of the 5´promoter region between -129 and -1 as here the sequence of the promoter regions of Ppact1 and Fhact1 is highly conserved as just mentioned and the -82 deletion construct does not contain all of the highly conserved sequence but the -237 deletion construct does.

Highly conserved regions within the promoters of Ppact3 and Fhact5 can also be observed. In this case the promoter regions for both genes isolated are much longer. Therefore even more regions of homologies are found between the two 5´promoter regions (Fig. 19 B). In this case the promoter regions of Ppact3 from -1 to -2270 and of Fhact5 from -64 to -2325 show some interesting homology features. The difference in the TS position might be due to the fact that the 5´UTR of Fhact5 was determined experimentally and the one of Ppact3 was determined by analysing ESTs from database.

The sequence of Ppact3 between -2270 and -1876 shows only a 29 % low homology to the same sequence area of Fhact5 located between -2325 and -1948. Then an expanded region of about 1100 nt is following showing a very high homology of 82 %. The next 140 nt of Ppact3 and 152 nt of Fhact5 promoter show "only" 53 % homology. The sequence of Ppact3 located between -641 and -463 shows again high conservation of 76 % to the region between -705 and

-528 of Fhact5. The following about 180 nt show again lower homology of 53 %. The last 288 bp of Ppact3 promoter sequence then are again more homolog with 73 % to the next 280 bp of Fhact5. These regions of different degrees of homologies between the two homologous genes might indicate the presence of regulative active elements within the 5´promoter region.

As for the case of Ppact1 and Fhact1 also here the expression analysis of the different Ppact3:vegf deletion constructs are interesting in this context (Fig. 17). Here a significant increase of the vegf expression level of the -2210, -995, -821 deletion constructs compared to the -523 deletion construct was observed. The three deletion construct which contain at least parts of the expanded homolog region between -1876 and -779 found in Ppact3 and Fhact5 reached levels about that of the 35 S promoter whereas the -523 deletion construct showed a 2 ½ fold increase of expression compared to the 35S promoter or the longer deletion constructs. This might argue for the presence of a negative regulator within this region of 82 % homology between Ppact3 and Fhact5.

In the case of Marchantia, no comparable sequence homologies could be found between the different actin genes from Physcomitrella and Funaria.

For the Fhact5 gene a construct containing 1157bp of the 5´untranscribed region fused to the hVEGF cDNA was made and used for transient transformation experiments on Physcomitrella protoplasts. The amount of protein detected in this case was in the same range but slightly higher (up to 2 folds) as with the CaMV 35S promoter. The Fhact5 gene presents the highest homology to the PpAct3 gene, and interestingly both of the promoters showed a similar activity in Physcomitrella protoplasts during the transient assays.

2.5.Stable transgenic lines.

The cassettes containing Ppact1, Ppact5 and Ppact7 5' MEPRs driving the expression of the VEGF cDNA were introduced in the genome of Physcomitrella plants. For each of the MEPRs five to

ten stably transformed plants were recovered and tested for the expression of rhVEGF. For these three MEPRs tested, expressed and secreted moss derived rhVEGF was detected in the supernatants of the cultures where the plants were growing (standard Knop medium), indicating that the MEPRs promote protein expression under non- inducing conditions (standard conditions) when they are integrated in other parts of the genome. The amount of protein that could be measured in those lines ranged from 7ngVEGF/mg moss dry weight until 53ngVEGF/mg moss dry weight, depending on the construct and the stable line.

One transgenic moss strain containing VEGF cDNA under control of Ppact5 was used to perform bioreactor cultures. The amount of moss derived recombinant VEGF in the supernatant of bioreactor cultures measured by ELISA was 40-50ngVEGF/mg moss dry weight.

EXAMPLE 3: Cloning and analysis of Physcomitrella patens and Funaria hygrometrica ubiquitin genes and their expression promoting regions.

Taking advantage of the presence of several EST sequences corresponding to polyubiquitin genes of Physcomitrella, specific oligos were designed to isolate the corresponding genomic sequences of the most abundantly present EST of the ubiquitin gene homologous sequence in the databases, named Ppubq1. 2146bp of 5'region of Ppubq1 could be identified by iPCR . A 129bp transcribed 5´leader is present before the ORF starts, determined by 5'race. The 5'region of Ppubq1 is amplified by PCR on genomic DNA from Physcomitrella patens using the primers 777 and 602.

Vectors carrying different parts of promoter and 5´UTR region driving expression of the hVEGF cDNA, were constructed to analyse the activity of the the promoter during transient transformation of Physcomitrella protoplasts.

The results indicated a similar activity for this promoter to the Ppact5 promoter (or even higher). The constructs tested, 1.6Kb and 1.3Kb promoter fragments, reached expression levels around 4 times and almost 7 times higher than the CaMV 35S.

The ubiquitin gene from Funaria, Fhubq1, was identified by performing a 5' race PCR on Funaria total RNA with a primer derived from the Ppubq1 coding sequence. The isolated 5'UTR sequence and partial coding sequence was used to design primers for iPCR on genomic ligations of *Funaria hygrometrica*. This way 5' upstream sequence of the 5'UTR was identified. The 5'region is amplified by PCR on genomic DNA from Funaria hygrometrica using the primers 943 and 944.

EXAMPLE 4: Cloning and analysis of Physcomitrella patens RBCS expression promoting regions.

As putative candidates next to the actin, tubulin and ubiquitin genes the ribulose-1,5- bisphospate carboxylase/ oxygenase small subunit (rbcS) genes were taken into consideration. The different rbcS genes are encoded on the nuclear genome. The rbcS genes are members of a gene family. The rbcS genes are expressed basically in all green parts of plants able to fixate CO_2. Therefore this gene family is of interest to get 5'and 3' flanking expression promoting regions of different rbcS genes from different mosses. As a first step Physcomitrella EST databases were analysed. It was found that the rbcS genes from Physcomitrella patens are organised in a gene family, consisting of 12 genes. The most abundantly present ESTs of the rbcS genes, named Pprbc-S12, was taken as a candidate to find it's 5' and 3' expression promoting sequences. Starting with the EST sequence data, 5' and 3' flanking regions of this gene was identified by iPCR and the cloned 5'and 3' regions were sequenced. The 5'region is amplified by PCR on genomic DNA from Physcomitrella patens using the primers 839 and 858. The 3'region is amplified by PCR using the primers 904 and 901.

In the enclosed Sequence Listing, the following sequences are given (Seq.ID.No/name of sequence/ 5' or 3' region relative to the protein encoding region):

1 Pptub1 5'
2 Pptub1 3'
3 Pptub2 5'
4 Pptub2 3'

5	Pptub3 5'
6	Pptub3 3'
7	Pptub4 5'
8	Pptub4 3'
9	Ppact1 5'
10	Ppact1 3'
11	Ppact3 5'
12	Ppact3 3'
13	Ppact5 5'
14	Ppact5 3'
15	Ppact7 5'
16	Ppact7 3'
17	Fhact1 5'
18	Fhact1 3'
19	Fhact4.4 5'
20	Fhact5 5'
21	Mpact1 5'
22	Mpact4 5'
23	Mpact15 5'
24	Ppubq1 5'
25	Fhubq1 5'
26	PprbcS12 5'
27	PprbcS12 3'

Table 1: List of primers

No.	sequence (5´-3´)
35	ATCCAGGAGATGTTCAGGCG
36	CCGMACGCTGTCCATRGTYCC
38	ACATTGATGCGCTCCARCTGC
40	GGBATGGACGAGATGGAGTTCAC
67	AGCACATGCACACCCAATACGCTTGTCGCAATTC
69	GTCGTCATAGACGACAAGACCGGGGATCCACAGC
70	TCAGTGCTGTCCGTGAATCTCTCTCTCTGCTTTG
71	CTGTGTTCGGATTAGACTCCCCGTAGCCTTTGTG
89	TCGATTGGCGAGTTGCGAAGGAGGGCAAGG
90	TGCCTGCTCATCTTGAGTATGGCGTGTTG
91	CTGCAAGCAATGCGCACTGAAACAAGATGG
95	GACCTGGAAACCTGCACAATCACGCATAGA
113	TAGCATAAGATAAAGATGTTCTCTACC
149	CTCACCAGCCAATGGCTATGC
203	CCGTGGGACTTAGTTGTCTTCACTTC
204	GATCGAAATTGCTGCTTGGCCTCCAC
205	TCGCAGGATGTGTCCTTAGTCGAGAA
206	AACTTCACGCATTCCACAAGCCACAC
219	TTGATACTCGAGAAGTCCAAAATAATTTAATGATAC
223	CATCTTCGCTAAGGATGATCTACAACGAG
225	CATCTTCAGTGTGCTCTACCTCACG
226	CTACTCGAGCACATATAATACTGCCCTAGTGCC
291	GACAGATCTCCTTAGTCGAGAAGGCGCGGGACGTG
292	GACCCGTGGGACTTAGTTGTCTTCACTTC
296	GCTGCTCTTCTCGTGATTGTCT
297	CATTCCCACCCTTCCTTCTCTTC
298	GTTTTGTGGCTCTTCCTTGG
299	ATCGCTTCTCGACTCTTCTTCC
300	GTTACGCTCGCAATGCGTACT
320	AACTTTCTGCTGTCTTGGGTGCATTG
321	GACCTGCAGGCACTCGAGCTTGTAATCATGGTCATAG
332	CATTTCTTAATACCGACCTGCCCAACCA
333	CATGGAGAAGAAATACTCTGCACATCAAAAG
334	CATTATTTAATACGGACCTGCACAACAAC
335	CATTTTTTAGAATGATCCTACAGGAGTTC
336	AGATCTGGCAAGTTCCCTTCG
337	GAAGAGAAGGAAGGGTGGGAATG
338	GGAAGAAGAGTCGAGAAGCGAT

No.	sequence (5´-3´)
363	CATCTTGTCCAACTACCGCGACCCGAACCC
364	AATCTCGAGTAGCATAAGATAAAGATGTTCTCTACC
373	GGTAAAGCTTCTCGAGTGCAGTAGACGACAAAATG
374	CATCTTGCTCAAGCTGTGCGAAGCTC
395	ATCTCGAGGATCCATTCAACGGAGGATAAGT
408	CAACTCGAGATCGGTCTGTAAGCCCTGTATTTG
413	ATTTCTCGAGTTGTTGAATCATGTTAATTGCCAATGGT
511	TTACTCGAGACTCTACTAATTGACAAGTATG
547	GTCAAGATTGGAGGTTCCTTGAG
548	TCCATCTCGAGTACCTCCGCTGTGTGTTTCAAAG
549	GTGCCTCGAGCCACATCCCGACCGCC
550	AGCACCTCGAGTACTGCCCTAGTGCCCTAATC
602	CATCCTTACAGGACGTACTGG
611	ATGCATGGCAAAACATCCCCTG
612	CATGGAGATGAAATGTTCTG
777	TTAACTCGAGATACAAGAGTTATAAATCATATAC
839	ATATCTCGAGATGCATGTAAGATAATTCCAATTAGA
858	CATTGCTAAAATCTCTCCACACTCGAATC
901	ATATCTGCAGTCATGAAACTTTCATTATGTATC
904	ATATGCGGCCGCGGAACGAATTTGTCGAGCTCTCT
908	CTTTCGTGTTGCCTCAAGAGTG
909	CATTTCTTAATACGGACCTGCC
943	ATATCTCGAGGAATTCATTTCCATTAACGAGAATATGAC
944	CATCTTCACAACGCTTTATCACTTC
950	CATATGCGTACGGAGTTGTGG
951	TTTCGCGAAGTTACCTAACC
960	TCATGATGTTAAGCGTTTTCA
961	GTTAACGAAGGAGGTGTCCG
970	AAGCTTAGCAAGCAGCTCTCGCAG
971	ATCGACGATAGACTGCAAGCC
983	AGGAGTGTTACACATCTTTTAC
984	GGCTAAGACGACGCATTCTGTG
1000	GGATCCGAGAGGAAAGAGAGAG
1001	CGCTTACAATGATCCTGCATAG
10R	TCDGTGAACTCCATCTCGTCCAT
8F	CGGTACCTACAAGGGCCTCTCG
9F	TGGGACGTATCAGGGTACGTCT
F7	TATCCGGAGGTTCCCGCGACACC

REFERENCES:

Altschul et al., Nucleic Acids Research 25 (1997), 3389-3402.
Berlin et al., Genetic modification of plant secondary metabolism: Alteration of product levels by overexpression of amino acid decarboxylases, in: Advances in Plant Biology, Studies in Plant Science, Vol. 4, pp 57-81, DdY Ryu and S Furasaki (eds), Elsevier, Amsterdam 1994.
Cove et al. Trends Plant Sci. 2 (1997), 99-105
Engel, Am. J. Bot. 55 (1968), 438-446
EP 1 206 561 A
Frahm "Biologie der Moose" (2001), Spektrum Akad. Verlag
Gorr (1999) Biotechnologische Nutzung von Physcomitrella patens (Hedw.) B.S.G.. Dissertation, Universität Hamburg.
Grimsley et al., Mol. Gen. Genet. 154 (1977), 97-100
Hiwatashi et al., Plant J. 28(1) (2001), 105-116
Hohe et al., Plant Sci. 163 (2002), 69-74
Holtdorf et al., Plant Cell Rep. 21 (2002), 341-346
Liebich et al., Nucleic Acids Research 30 (2002), 3433-3442.
McElroy et al., Mol. Gen. Genet. 231 (1991), 150-160
Reski et al., Planta 165 (1985), 354-358
Reski et al., Mol. Gen. Genet. 244 (1994), 352-359
Reski, Botanica Acta III (1998), 1-15
Reski, Planta 208 (1999), 301-309
Reutter et al., Plant Tissue Culture and Biotechnology 2 (1996), 142-147
Richter et al., Gene 290 (2002), 95-105
Rothnie et al., Plant Mol Biol.32 (1996), 43-61.
Rother et al., J. Plant Physiol. 143 (1994), 72-77
Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press
Schaefer et al., Mol Gen Genet 226 (1991), 418-424
Schaefer (2001) Principles and protocols for the moss Physcomitrella patens. http://www.unil.ch/lpc/docs/PPprotocols2001.pdf
Schaefer, Annu. Rev. Plant Biol. 53 (2002), 477-501
Schlink et al., Plant Mol. Biol. Rep. 20 (2002), 423a-423f
Strepp et al., Proc. Natl. Acad. Sci. USA 95 (1998), 4368-4373
Töpfer et al., NAR 15 (1987), 5890
Watson et al., "Recombinant DNA" (1992), Chapter I.1 and 2
Zeidler et al., J. Plant Physiol. 154 (1999), 641-650